Linear Regression: Equations

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Objective

- 데이터셋($\mathcal{D}$)이 주어졌을 때, loss를 최소로 하는 파라미터(θ)를 찾자

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N},$$
$$\text{where } x_{1:N} \in \mathbb{R}^{N\times n} \text{ and } y_{1:N} \in \mathbb{R}^{N\times m}.$$

$$x_{1:N} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,n} \end{bmatrix}$$

$$\hat{\theta} = \underset{\theta\in\Theta}{\text{argmin}} \sum_{i=1}^{N} \|y_i - f_\theta(x_i)\|_2^2,$$
$$\text{where } \theta = \{W, b\} \text{ and } f_\theta(x) = x \cdot W + b.$$

$$y_{1:N} = \begin{bmatrix} y_{1,1} & \cdots & y_{1,m} \\ \vdots & \ddots & \vdots \\ y_{N,1} & \cdots & y_{N,m} \end{bmatrix}$$

Loss Minimization using Gradient Descent

- Loss 함수를 파라미터(W, b)로 미분하여, 기울기 값을 활용해보자

$$\hat{y}_{1:N} = x_{1:N} \cdot W + b,$$
$$\text{where } W \in \mathbb{R}^{n \times m} \text{ and } b \in \mathbb{R}^{m}.$$

$$\begin{aligned}
\mathcal{L}(\theta) &= \sum_{i=1}^{N} \|y_i - \hat{y}_i\|_2^2 \\
&= \sum_{i=1}^{N} \sum_{j=1}^{m} (y_{i,j} - \hat{y}_{i,j})^2 \\
&= \sum_{i=1}^{N} \sum_{j=1}^{m} y_{i,j}^2 - 2y_{i,j} \cdot \hat{y}_{i,j} + \hat{y}_{i,j}^2,
\end{aligned}$$

$$W = \begin{bmatrix} w_{1,1} & \cdots & w_{1,m} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,m} \end{bmatrix}$$

$$\text{where } \hat{y}_{i,j} = \sum_{k=1}^{n} x_{i,k} \times W_{k,j} + b_j.$$

Loss Minimization using Gradient Descent – Detail

- Loss 함수를 파라미터(θ)로 미분하여, 기울기 값을 활용해보자

$$\theta = \{W, b\}, \text{where } W \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^{m} \qquad W = \begin{bmatrix} w_{1,1} & \cdots & w_{1,m} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,m} \end{bmatrix}$$

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{i=1}^{N} \|y_i - \hat{y}_i\|_2^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{m} (y_{i,j} - \hat{y}_{i,j})^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{m} y_{i,j}^2 - 2y_{i,j}\cdot\hat{y}_{i,j} + \hat{y}_{i,j}^2, \end{aligned}$$

$$\text{where } \hat{y}_{i,j} = \sum_{k=1}^{n} x_{i,k} \times W_{k,j} + b_j.$$

$$\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}(\theta)$$

$$\downarrow$$

$$W \leftarrow W - \eta\nabla_W\mathcal{L}(\theta)$$

$$b \leftarrow b - \eta\nabla_b\mathcal{L}(\theta)$$

$$\downarrow$$

$$W_{k,j} \leftarrow W_{k,j} - \eta\frac{d\mathcal{L}(\theta)}{dW_{k,j}}$$

$$b_j \leftarrow b_j - \eta\frac{d\mathcal{L}(\theta)}{db_j}$$